

The observational roots of reference of the semantic web^{*}

Simon Scheider¹, Krzysztof Janowicz¹ and Benjamin Adams²

¹ Geography, University of California, Santa Barbara simonscheider@web.de
jano@geog.ucsb.edu

² Computer Science, University of California, Santa Barbara badams@cs.ucsb.edu

Abstract. Shared reference is an essential aspect of meaning. It is also indispensable for the semantic web, since it enables to weave the global graph, i.e., it allows different users to contribute to an identical referent. For example, an essential kind of referent is a geographic place, to which users may contribute observations. We argue for a human-centric, operational approach towards reference, based on respective human competences. These competences encompass perceptual, cognitive as well as technical ones, and together they allow humans to inter-subjectively refer to a phenomenon in their environment. The technology stack of the semantic web should be extended by such operations. This would allow establishing new kinds of observation-based reference systems that help constrain and integrate the semantic web bottom-up.

1 Meaning and Reference

In the tradition of sciences such as philosophy and linguistics, semantics is the study of meaning, whereas syntax is the study of symbol manipulation. The classical distinction reflects that the realm of meaning is different and should not be confounded with the realm of syntax.

The things symbols stand for are called *referents*. The notion of meaning involves such essential problems as how humans come to *share* language referents, and how we can *refer* to them in an inter-subjective way. This problem is what we call the *problem of reference*. In many respects, it can be considered a practical manifestation of the *symbol grounding problem* [1], the problem of how to get out of the realm of syntax into any determinable contact with referents in the world. The problem of reference has been an ongoing issue in the semantic web [2], and symbol grounding recently was recognized as an untackled problem [3]. Like symbol grounding, the problem of reference is not solved by formal semantics of a Tarskian flavour, because formal theories can neither distinguish a particular domain nor a particular interpretation [4]. Furthermore, natural language descriptions cannot account for reference, too, because they are context dependent, in particular since they rely on polysemy and synonymy [4].

The semantic web was proposed to make the meaning of information on the web explicit. Reference is an essential part of meaning that is also indispensable for the semantic web, especially for linked data. In particular, it is shared reference which allows to establish identity, and, thus, allows different users to contribute information to the same information item, e.g., different tags to the same place in Linked Geo Data. This so called *network effect* is critical to the semantic web [5]. However, often, semantic web research is driven by syntax and technology stacks in the first place.

^{*} Dagstuhl Seminar 12221: Cognitive Approaches for the Semantic Web, 28.05.12 – 01.06.12, Dedre Gentner, Pascal Hitzler, Kai-Uwe Kühnberger, Frank van Harmelen

We argue for enlarging the breadth of semantic web research by focusing on the problem of reference. This requires us to learn from existing grounded symbol systems and to notice how humans cognize [6] and disambiguate referents in language.

2 An operational view on the roots of reference

We argue for an operational view on the problem of reference and take up on the ongoing debate in the semantic web.

Meaning and referencing cannot be reduced to symbols. Some authors have tried to account for reference in terms of URIs [7], formal semantics [8,9], and appropriate descriptions [5]. Such approaches rely on the assumption that while meaning is external to the web, it can be regarded as internal to the semantic web; see also [8,7]. However, it is important to note that while the language of the semantic web is machine readable, meanings are not. Hayes points out that it is not possible to solve the problem of reference within language constructs [2]. The tendency seems therefore to surrender and start praising the benefits of ambiguity in language instead [2,8]. The existence of ontological disagreement [8], however, does not mean that we can dispense with intended meaning, as Ginsberg argued [10]. Intended meanings provide reference, and shared reference is critically indispensable in order to prevent data isolation.

Referencing is something human observers do. We argue that there is no need to surrender but a need to shift perspective on the problem. Hayes and Halpin hint at the right direction when they state that “reference is part of a communicative act between cognitive agents” that involves conventions [2]. We argue that referencing (and meaning as well) is not an object or some abstract mapping, but something that humans do [11], namely a kind of *speech act* that draws human attention to some reproducible phenomenon [12]. However, one should be aware that understanding this act draws on almost the full range of human cognitive capabilities¹, including Gestalt perception, joint attention [13], conventionally established predications² and imagination.

Technical language reference is based on human operations. The problem of where to begin the construction of a concise technical language [14], i.e., where to ground terms [12], is decisive for its meaningfulness, and is much too often disregarded in information ontology. [14] have argued that in order to prevent talking at cross purposes, it is necessary to begin with a language level whose interpretation is internalized, embodied in practice, and therefore shared among speakers, and on which a more concise technical language can be constructed. Paul Lorenzen argued that in order to understand the method of thought, we need to comprehend the extra-linguistic actions that are needed to build a language [14].

Does this mean that the current semantic web technology stack is useless? No, since ontologies are useful to constrain the interpretation and to infer implicit knowledge. However, they constrain interpretation only top-down, while shared references are based on human capabilities that constrain interpretation bottom-up [11]. We argue that the *technology stack of the semantic web needs to draw on these cognitive capabilities* in order to allow bottom-up semantic constraints [16, c.f.].

Based on this argumentation, we have shown in previous work how to reconstruct *geospatial concepts* based on perceptual operations and constructions instead of describing them top-down [12,17]. For example, geographic places [18], water bodies and environmental media [19], environmental qualities [20], road networks and

¹ See the discussion in [12].

² See [14]. Quine’s *observation sentences* [15] reflect a similar idea.

junctions can be reconstructed based on observing *action potentials* (affordances) [21].

Reference systems. Solutions to the problem of reference should transgress syntax as well as technology. They cannot solely rely on computers but must also rely on human referential competences. This requirement is met by *reference systems* [22]. Reference systems are different from ontologies in that they constrain meaning bottom-up [11]. Most importantly, they are not "yet another chimaera" invented by ontology engineers, but already exist in various successful variants. A prominent example are *spatial (geodetic) reference system*, which provide a language about points on a mathematical ellipsoid anchored relative to the earth surface by reference observations of places and directions. It is responsible for the success of geographic information systems (GIS), navigation systems, and serves as an integration layer for the semantic web [23]. Another example are *measurement scales*, whose units of measurement are likewise established by conventional observations. *Address systems* are (geo-)reference systems for certain kinds of observable places, namely buildings. *Calendars* are temporal reference systems anchored in the observation of day and night, as well as celestial bodies.

New kinds of reference systems as semantic integration layer. A reference system can be seen as a theory in which abstract or vague concepts are reconstructed based on conventionally established and, therefore, reproducible observation procedures [12]. This makes it useful as a semantic plane for linking heterogeneous conceptualizations that still divide the semantic web. A comparable idea is Peter Gärdenfors' notion of a conceptual space [24]. This is a geometric, topological space whose dimensions are anchored in human cognition, and in which concepts are expressed as convex regions. For example, local conceptualizations of a term like "high mountain" can be compared in terms of different convex regions in this space [17]. New kinds of reference systems would allow us to negotiate or translate between heterogeneous views on the world [23]. For example, places are essential concepts for spatial referencing of user generated content on the social web, such as points of interest (POI), geo-tagged pictures, blog posts or tweets. But they are highly vague both regarding their boundaries as well as regarding category labels [12]. A *place reference system* would allow us to disambiguate place concepts beyond spatial footprints, taking into account the different types of afforded actions that constitutes them as well as their spatio-temporal distribution [18].

3 Conclusion

We argued that the technology stack of the semantic web should additionally draw on human operations that account for reference. This would enable us to establish and reconstruct reference systems, beyond the existing ones for measurements, locations, times and buildings. What we need are more and other kinds of reference systems in order to account for the various semantic domains involved in the semantic web. These may allow us, for example, to construct concepts as abstractions in terms of logical reifications or convex regions in conceptual spaces. They assure inter-subjectivity of reference based on joint human observation procedures and convention [12], but require us to broaden the focus of semantic research from top-down, language-centric to bottom-up, human-centric and operational approaches.

References

1. Harnad, S.: The symbol grounding problem. *Physica D* **42** (1990) 335–346

2. Hayes, P.J., Halpin, H.: In defense of ambiguity. *Int. J. Semantic Web Inf. Syst.* **4**(2) (2008) 1–18
3. Cregan, A.: Symbol grounding for the semantic web. In: *ESWC'07*. (2007) 429–442
4. Scheider, S.: The case for grounding databases. In: *GeoSpatial Semantics. Third International Conference, GeoS 2009, Mexico City. LNCS.* Springer, Berlin (2009) 44–62
5. Booth, D.: URIs and the Myth of Resource Identity. (2006)
6. Raubal, M., Adams, B.: The semantic web needs more cognition. *Semantic Web* **1**(1-2) (2010) 69–74
7. Berners-Lee, T., Fischetti, M.: *Weaving the web: The original design and ultimate destiny of the world wide web by its inventor.* Harper, San Francisco (1999)
8. Parsia, B., Patel-Schneider, P.F.: Meaning and the semantic web. In: *WWW (Alternate Track Papers and Posters)'04*. (2004) 306–307
9. Presutti, V., Gangemi, A.: Identity of resources and entities on the web. *Int. J. Semantic Web Inf. Syst.* (2008) 49–72
10. Ginsberg, A.: Ontological indeterminacy and the semantic web. *Int. J. Semantic Web Inf. Syst.* (2008) 19–48
11. Kuhn, W.: Semantic engineering. In Navratil, G., ed.: *Research Trends in Geographic Information Science. Volume 12 of Lecture Notes in Geoinformation and Cartography.* Springer, Berlin (2009) 63–76
12. Scheider, S.: *Grounding geographic information in perceptual operations.* PhD thesis, University of Münster (2011)
13. Tomasello, M.: *The cultural origins of human cognition.* Harvard University Press, Cambridge, MA (1999)
14. Kamlah, W., Lorenzen, P.: *Logische Propädeutik. Vorschule des vernünftigen Redens.* 3rd edn. J.B. Metzler, Stuttgart, Weimar (1996)
15. Quine, W.: *The roots of reference.* Open Court Publishing, La Salle, Illinois (1974)
16. Janowicz, K.: Observation-driven geo-ontology engineering. *Transactions in GIS* **16**(3) (2012) 351–374
17. Adams, B., Janowicz, K.: Constructing geo-ontologies by reification of observation data. In: *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. GIS '11, New York, NY, USA, ACM* (2011) 309–318
18. Scheider, S., Janowicz, K.: Places as media of containment. In: *Extended Abstracts Volume, GIScience 2010.* University Zürich-Irchel, Zürich (2010)
19. Scheider, S., Janowicz, K., Kuhn, W.: Grounding geographic categories in the meaningful environment. In Hornsby, K., Claramunt, C., M., D., Ligozat, G., eds.: *Spatial Information Theory, 9th International Conference, COSIT 2009, Aber Wrac'h, France, September 21-25, 2009 Proceedings.* Springer, Berlin (2009) 69–87
20. Scheider, S., Keßler, C., Ortmann, J., Devaraju, A., Trame, J., Kauppinen, T., Kuhn, W.: Semantic referencing of geosensor data and volunteered geographic information. In Ashish, N., Sheth, A., eds.: *Geospatial Semantics and the Semantic Web. Foundations, Algorithms, and Applications. Volume 12 of Semantic Web and Beyond.* Springer, Berlin (2011) 27–59
21. Scheider, S., Kuhn, W.: Affordance-based categorization of road network data using a grounded theory of channel networks. *International Journal of Geographical Information Science* **24**(8) (2010) 1249 – 1267
22. Kuhn, W.: Semantic reference systems. *Int. J. Geogr. Inf. Science* **17** (5) (2003) 405–409
23. Janowicz, K.: The role of space and time for knowledge organization on the semantic web. *Semantic Web Journal* **1**(1-2) (2010) 25–32
24. Gärdenfors, P.: Conceptual spaces as a framework for knowledge representation. *Mind and Matter* **2**(2) (2004) 9–27